A data-independent acquisition (DIA) assay library for quantitation of environmental effects on the kidney proteome of *Oreochromis niloticus*

Larken Root[1], Aurora Campo[2], Leah MacNiven[1], Pazit Con[2], Avner Cnaani[2], and Dietmar Kültz[1]

[1]University of California Davis
[2]Agricultural Research Organization Volcani Center

November 16, 2020

Abstract

Interactions of organisms with their environment are complex and environmental regulation at different levels of biological organization is often non-linear. Therefore, the genotype to phenotype continuum requires study at multiple levels of organization. While studies of transcriptome regulation are now common for many species, quantitative studies of environmental effects on proteomes are needed. Here we report the generation of a data-independent acquisition (DIA) assay library that enables simultaneous targeted proteomics of thousands of *Oreochromis niloticus* kidney proteins using a label- and gel-free workflow that is well suited for ecologically relevant field samples. We demonstrate the usefulness of this DIA assay library by discerning environmental effects on the kidney proteome of *O. niloticus*. Moreover, we demonstrate that the DIA assay library approach generates data that are complimentary rather than redundant to transcriptomics data. Transcript and protein abundance differences in kidneys of tilapia acclimated to freshwater and brackish water (25 g/kg) were correlated for 2114 unique genes. A high degree of non-linearity in salinity-dependent regulation of transcriptomes and proteomes was revealed suggesting that the regulation of *O. niloticus* renal function by environmental salinity relies heavily on post-transcriptional mechanisms. The application of functional enrichment analyses using STRING and KEGG to DIA assay datasets is demonstrated by identifying *myo*-inositol metabolism, antioxidant and xenobiotic functions, and signaling mechanisms as key elements controlled by salinity in tilapia kidneys. The DIA assay library resource presented here can be adopted for other tissues and other organisms to study proteome dynamics during changing ecological contexts.

Title: A data-independent acquisition (DIA) assay library for quantitation of environmental effects on the kidney proteome of *Oreochromis niloticus*.*Running Title*: Tilapia quantitative DIA proteomics libraries*Authors*: Larken Root[1], Aurora Campo[2], Leah MacNiven[1], Pazit Con[2], Avner Cnaani[2] and Dietmar Kultz[1]*Author affiliations*:[1]Department of Animal Sciences, University of California Davis, Meyer Hall, One Shields 10 Avenue, Davis, CA 95616, USA[2]Department of Poultry and Aquaculture, Institute of Animal Sciences, Agricultural Research Organization, Volcani Center, P.O. Box 15159, Rishon LeZion 7528809 Israel

Abstract :

Interactions of organisms with their environment are complex and environmental regulation at different levels of biological organization is often non-linear. Therefore, the genotype to phenotype continuum requires study at multiple levels of organization. While studies of transcriptome regulation are now common for many species, quantitative studies of environmental effects on proteomes are needed. Here we report the generation of a data-independent acquisition (DIA) assay library that enables simultaneous targeted proteomics of

thousands of *Oreochromis niloticus* kidney proteins using a label- and gel-free workflow that is well suited for ecologically relevant field samples. We demonstrate the usefulness of this DIA assay library by discerning environmental effects on the kidney proteome of *O. niloticus* . Moreover, we demonstrate that the DIA assay library approach generates data that are complimentary rather than redundant to transcriptomics data. Transcript and protein abundance differences in kidneys of tilapia acclimated to freshwater and brackish water (25 g/kg) were correlated for 2114 unique genes. A high degree of non-linearity in salinity-dependent regulation of transcriptomes and proteomes was revealed suggesting that the regulation of *O. niloticus* renal function by environmental salinity relies heavily on post-transcriptional mechanisms. The application of functional enrichment analyses using STRING and KEGG to DIA assay datasets is demonstrated by identifying *myo* -inositol metabolism, antioxidant and xenobiotic functions, and signaling mechanisms as key elements controlled by salinity in tilapia kidneys. The DIA assay library resource presented here can be adopted for other tissues and other organisms to study proteome dynamics during changing ecological contexts.

Keywords: Adaptation, Aquaculture, Quantitative proteomics, Teleost

Introduction:

Explaining emergent properties of complex organisms represents a grand challenge of biology (Schwenk, Padilla, Bakken, & Full, 2009). Emergent properties are often based on non-linear interactions of their molecular constituents (Stillman et al., 2011). Despite the known non-linearity of the genome to phenome continuum many biological studies rely heavily on correlations of complex organismal phenotypes (traits) with genomic variation (e.g. SNPs and other types of sequence variation) via QTL or GWAS analyses (Kratochwil & Meyer, 2015; Wray et al., 2013) and mRNA abundance changes via transcriptomics while proteome dynamics is less commonly investigated (Evans, 2015).

Unlike the genome, transcriptomes and proteomes are spatially highly heterogeneous and dynamic, i.e. they are highly responsive to environmental stimuli. Currently, there is a large gap in the literature regarding transcriptome versus proteome data and systematic comparisons aimed at discerning the rules of non-linearity between these two levels of biological organization that are not common (Buccitelli & Selbach, 2020). There is a great need for developing robust and comprehensive quantitative proteomics approaches to facilitate such comparisons and close the gap between genotypes and ecologically relevant phenotypes, e.g. environmental stress tolerance. Understanding the connection between mRNA and protein levels is fundamental to predicting how the underlying genetic code impacts changes in phenotype due to environmental changes. Non-linearity between transcriptome and proteome levels of regulation is well documented (Franks, Airoldi, & Slavov, 2017). It can be based on differential mRNA processing and degradation (Bentley, 2014), transcript-specific regulation of protein translation through all stages including initiation, elongation, and localization (Tahmasebi, Khoutorsky, Mathews, & Sonenberg, 2018), and/ or regulation of protein degradation (Pohl & Dikic, 2019).

The proteome represents the core that is central to the genome to phenome continuum. On the one hand, proteins are linked directly to specific genes via proteotypic peptides, which allows unambiguous association of each protein with a specific genomic locus (Keerthikumar & Mathivanan, 2017). On the other, proteins represent the critical molecular building blocks that define structure and carry out most biochemical processes and functions of cells, tissues and organisms (Ebhardt, Root, Sander, & Aebersold, 2015). Proteins represent the molecular constituents giving rise to phenotypic variability that is acted upon by natural selection (Clarke, 1971; Frömmel & Holzhütter, 1985; Mularoni, Ledda, Toll-Riera, & Albà, 2010). Moreover, most targets of pharmaceutical drugs are proteins, which illustrates that proteins control critical organismal phenotypes (Batchelor, Loewer, Mock, & Lahav, 2011; Ebhardt et al., 2015).

Recent developments in biological mass spectrometry have enabled robust gel- and label-free quantitative proteomics workflows that are well suited for organismal biology and molecular ecology (Huang et al., 2015; Vowinckel et al., 2013). In particular, the invention of data-independent acquisition (DIA) liquid chromatography mass spectrometry (DIA-LCMS2) holds great promise for molecular ecology studies (Crowgey,

2

Matlock, Venkatraman, Fert-Bober, & Van Eyk, 2017; Schubert et al., 2015). The DIA approach is also referred to as Sequentially Windowed Acquisition of all theoretically possible MSMS spectra (SWATH)-MS (Arnhard, Gottschall, Pitterl, & Oberacher, 2015; Huang et al., 2015).

DIA-LCMS2 represents a merger of pre-acquisition targeted mass spectrometry approaches, i.e. selected reaction monitoring (SRM) or multiple reaction monitoring (MRM), and non-targeted data acquisition that is independent of precursor (MS1) spectra acquisition (Koopmans, Ho, Smit, & Li, 2018). In DIA-LCMS2 the targeting of specific transitions, precursors, peptides, and proteins is performed post-acquisition by interrogating all theoretically possible fragment ion (MS2 spectra) present in a sample against a previously validated DIA assay library. Here, we have constructed a DIA assay library from raw MS2 spectral libraries of Nile tilapia (*Oreochromis niloticus*) kidney to facilitate quantitative studies of proteome dynamics in response to environmental stress and other ecological contexts. We demonstrate the utility of this DIA assay library by identifying proteins, biological functions, and processes that are associated with salinity acclimation in *O. niloticus* kidney. Furthermore, we demonstrate that quantitative proteomics provides knowledge that cannot be gained from transcriptomics data.

Methods :

Origin of samples for DIA assay library construction and RNA sequencing

The *O. niloticus* stock used in this study originated from Lake Manzala, Egypt, and brought to Israel in the 1970's. Twelve male fish (159±33 g) were randomly distributed between two 600L tanks. After two weeks acclimation, one group was exposed to a gradual salinity increase of 5 g/kg per day up to a final salinity of 25 g/kg using synthetic sea salt (Red Sea Salt, Eilat, Israel). Fish were fed twice per day, 2% of their body weight (feeding was ceased 24 hours before sample collection). After 24 hours at the final salinity, fish were sacrificed and immediately dissected. Kidney samples were removed and divided into two aliquots, which were immediately snap-frozen in liquid nitrogen and stored at -80°C. One aliquot from each individual was subsequently processed for proteomics, the other was processed for RNAseq.

Sample processing for quantitative proteomics

Extraction and in-solution trypsin digestion of proteins were performed as previously reported (Kültz, Li, Gardell, & Sacchi, 2013) with the following changes: After transfer of crushed sample powder to a low retention Eppendorf tube (LR-MCF), proteins were dissolved in 8M urea at 1.5x sample wt/vol and reduced with 10mM dithiothreitol (DTT) for 30 min. at 37°C. Proteins were then alkylated with 30 mM iodoacetamide (IAA) at RT for 30 min in the dark. Proteins were precipitated in an ice-cold solution of 10% trichloroacetic acid (TCA)/90% Acetone/0.15% DTT (5x the total volume). Samples were then centrifuged at 20,000 g for 5 min (4 °C) and the precipitated protein pellet washed once in 100% acetone/0.15% DTT. The protein pellet was re-dissolved in 8M urea (5x the original tissue wt/vol) for 30 min at RT on a rotator. After centrifugation at 19,000 g (5 min,) the supernatant was transferred to a clean LR-MCF tube and stored at 4 °C. Duplicate aliquots were used to determine protein concentration using a BCA protein assay compatible with diluted urea (Thermo-Pierce, cat. 23225). The appropriate amounts of LCMS grade water and 1M ammonium bicarbonate (Ambic) buffer (pH 8.5, final concentration 100 mM) added to dilute samples to 150 ng/100 µl total protein concentration in 0.6 ml LR-MCF tubes. Immobilized trypsin (Promega cat. V9012) was added at a 1:25 ratio relative to total protein and the samples incubated in a rotator at 35 °C for exactly 16h. Trypsin beads were removed by centrifugation for 2 min at 500 g and supernatant was transferred to a clean 0.6 ml LR-MCF tube. Any remaining undissolved solids were removed by centrifugation for 5 min at 19,000 g and supernatant was transferred to another clean 0.6 ml LR-MCF tube. Samples were then dried by speedvac (Thermo-Savant, ISS-110) until urea precipitate began forming. Peptides where then resuspended in 150 µl LCMS grade water containing 0.1% formic acid (FA) and transferred to total recovery glass vials (Waters 186000384C) for sample injection.

Peptide samples (2 µl, 100 ng/µl) were injected with a nanoAcquity sample manager (Waters, Milford, MA), trapped for 1 min at 15 µL/min on a Symmetry trap column (Waters 186003514), and separated on a 1.7µm particle size BEH C18 column (250mm x 75µm, Waters 186003545) by reversed phase liquid chromatography

using a nanoAcquity binary solvent manager (Waters). Peptides were eluted using a 125 min linear gradient ranging from 3% to 35% acetonitrile (ACN) directly (online) into a UHR-qTOF mass spectrometer (Impact II, Bruker) using a pico-emitter tip (New Objective FS360–20-10-D-20, Woburn, MA). Batch processing of samples was controlled with Hystar 4.1 (Bruker) and a quality control standard (68 fmol BSA peptide mix) was used at least once a week to monitor instrument performance.

Data dependent acquisition (DDA) for raw spectral library construction

Peak lists were generated from DDA raw data with DataAnalysis 4.4 (Bruker Daltonics) and peptide to spectrum matches identified with PEAKS suite X plus (Bioinformatics Solutions Inc., Waterloo, Canada) and X!Tandem Alanine (The GPM) database search engines followed by unambiguous assignment of peptides to unique proteins using the *O. niloticus* proteome database downloaded from NCBI RefSeq on Feb. 25, 2020. The database searched included 61,681 proteins plus the same number of randomly scrambled decoys and 282 common contaminants (human keratins, porcine trypsin, etc.). Trypsin cleavage was specified with at the C-terminus of either Lys or Arg except when followed by Pro, and two missed cleavages were the allowed maximum. Further, possible PTMs were searched first by allowing for Cys carbamidomethylation, Met oxidation, and Protein N-terminal acetylation. A second-round search considered all 313 PEAKSPTM modifications included in the default PEAKS suite X plus database, allowing max. 3 PTMs per peptide. Mass tolerance limits were set at 10 ppm for precursors and 0.03 Da for fragment ions. All DDA data and metadata are available at MassIVE (MSV000085637, doi:10.25345/C5DQ74) and ProteomeXchange (PXD020056).

Raw spectral library and DIA assay library construction

An *O. niloticus* kidney spectral library was generated using the peptide-to-spectrum matches and protein annotations generated from DDA data. Data was exported from PEAKS suite X plus in pepxml format and imported into Skyline 20.0 (Pino et al., 2017) to create a non-redundant raw library of MS2 spectra. The initial target list of proteins was then filtered using multiple QC criteria (see results) to reduce the number of transitions, precursors, peptides, and proteins to a unique set that has the highest diagnostic value for quantitation. The final filter steps utilized a sample training set, which was acquired by DIA as described in the next paragraph. The complete assay library including all relevant metadata and corresponding data for the sample training set is available at Panorama Public (Kueltz-Lab_2020-1.url). This assay represents a tier two assay (Abbatiello et al., 2017).

Data Independent Acquisition (DIA)

Each sample was analyzed by a second acquisition in data independent (DIA) mode. LC separation parameters and conditions were identical to those used for DDA but only MS2 spectra were acquired. The mass range for DIA was set to 390 – 1015 m/z at 25 Hz scan rate with an isolation width of 10 m/z (1 m/z overlap, 2.5 sec scan interval). Quantitative analyses and visualization of DIA data was performed using Skyline 20.0 (Pino et al., 2017). At least four (generally six) transition peaks were detected for each peptide and scored using the mProphet algorithm integrated into Skyline. The mass error threshold was set at 20 ppm for transitions, and the resolving power was 30,000. Randomly scrambled decoys were used in mProphet Q-value calculation. The R package MSstats 3.1 (Choi et al., 2014) was used for power analysis to calculate the fold-change (FC) cutoff that is appropriate for this experiment, as well as statistical significance of differences between treatment and control. The cutoff for multiple testing adjusted p-values (Benjamini & Hochberg, 1995) was set to $p<0.05$. MSstats analyses were normalized by equalizing medians at a minimum confidence interval of 95% with protein quantity as the scope for each analysis. Tukey's median polish was used as the summary method for MSstats analyses to weight each transition and each peptide of a given protein equally and the minimal mProphet detection Q-value for peak quality was set to 0.01.

Sample processing for RNAseq and quantitative analysis of transcriptomes

mRNA was extracted using TRIzol? reagent (Thermo Fisher Scientific, Waltham, MA), and purified to remove DNA contamination using the TURBO DNA-freeTM kit (Invitrogen, Carlsbad, CA). mRNA sample

4

were sent to the Israel National Center for Personalized Medicine (INCPM) at the Weizmann Institute of Science (Rehovot, Israel), where quality was determined on TapeStation Agilent 2200 system before library preparation and sequencing on an Illumina Hi-Seq 2500 device. Raw sequencing single ended data files were retrieved for trimming and adapter removal using Trimmomatic (v. 0.39, Bolger et al. 2014). After passing quality control assessment with FASTQC software (v. 0.11.9, Andrews et al. 2010), the fastq files were mapped against the *O. niloticus* reference genome (GCF_001858045.2) using STAR software (v. 2.7.3a, Dobin et al. 2013). The resulting BAM files from the mapping were processed with the package HTSeq (v. 0.11.1, Anders et al. 2014) in R (v. 3.6.3, R Core Team 2020) for obtaining the gene counts which were then submitted to the DEseq package (v. 1.39.0, Anders et al. 2010) in R (v. 3.6.3, R Core Team 2020) for retrieving up- and down-regulated genes, according to the experimental design. Significance was considered after adjusted p-value less than 0.05. Significant differentially expressed genes (DEG) were evaluated for their gene ontology (GO) annotations. All transcriptomics data are available at NCBI (https://submit.ncbi.nlm.nih.gov/subs/biosample/SUB8325839/).

Comparison of protein and transcript regulation

Mean protein abundances per treatment group were compared for each unique accession number with the corresponding mean mRNA abundance. Fold change (FC), the abundance difference between treatments expressed in multiples of the control, was transformed by the log2 and plotted with mRNA abundance along the X-axis as the independent variable and protein abundance along the Y-axis as the dependent variable based on mRNA-protein correlation analysis convention (Buccitelli & Selbach, 2020). Additionally, significantly regulated (adj. $p<0.05$) proteins were compared as a subset in the same manner, with the degree of correlation in each case being the coefficient of determination (R^2).

Genesis clustering

Protein regulation patterns were clustered in a non-biased manner using Genesis 1.8.1 (Sturn, Quackenbush, & Trajanoski, 2002). Proteins with similar regulation patterns in individual replicates were clustered to identify non-significant proteins which nonetheless followed the same regulation patterns as significant proteins. Six clusters were determined to be the ideal number which grouped the majority of all significantly regulated proteins in two groups, one containing most up-regulated proteins and the other containing most down-regulated, as well as isolating the majority of proteins in clusters with regulation patterns that did not follow treatment groupings.

STRING and KEGG analyses

Protein-protein interaction networks were analyzed using STRING ver. 11.0 (Szklarczyk et al., 2019) (https://string-db.org/). A list of protein accession numbers consisting of all significantly regulated proteins and mRNA, as well as those found in the up- and down-regulated protein clusters based on Genesis output was entered into the STRING search function. Search results created lists of significantly enriched (FDR<0.05) Uniprot keywords and protein domains classified by Pfam, InterPro, and SMART databases. Further, protein networks with more than one edge and containing at least one significantly regulated protein were identified and searched as a subset of the total group. From this result, in addition to the enriched keywords and domains, sub-networks where identified. An MCL inflation factor of 1.3 was chosen to create seven distinct networks which encompassed all of the significantly regulated proteins connected to at least one other protein. Each individual network protein list was then entered as a STRING search to determine enriched keywords and proteins and classify networks by cellular function.

Over-representation of proteins in known molecular pathways was analyzed using KEGG Mapper search (https://www.genome.jp/kegg/tool/map_pathway1.html). First, the GhostKOALA automatic annotation server (Kanehisa, Sato, & Morishima, 2016) was used to determine matching KEGG numbers to all of the proteins in the dataset. The complete list of matching KEGG numbers was searched in KEGG mapper to establish a baseline for representation for each pathway in the database. Then, lists of significantly up- and down-regulated proteins were searched, and the proportion of representation in pathways (number of proteins in a given pathway divided by total proteins in the list) compared against proportion for the complete

5

DIA assay library protein list. This approach indicated pathways with over-representation of significantly regulated proteins.

Results

Filtering of the raw spectral library and final composition the DIA assay library

We generated a raw MS2 spectral library from DDA data of all twelve samples, which contained 7004 proteins, 82605 peptides, 90350 precursors (different charge states of peptides), and 462868 transitions (fragment ions) (Figure 1). Transitions for the initial target list were chosen automatically from library spectra using Skyline transition settings based on the following criteria: ion 3 to last ion -1; fragment ion charge 1; precursor charge range $1 - 5$; mass accuracy threshold within 20 ppm of the expected mass. This initial target list was filtered in seven sequential steps using Skyline following a previously published method (Li, Levitan, Gomez-Jimenez & Kültz, 2018). Briefly, steps included (1) requiring a minimum number of 2 peptides per protein and 4 transitions per precursor, (2) elimination of peptides with more than 1 missed cleavage, (3) exclusion of PTMs other than Met/ Pro oxidation, Cys carbamido-methylation, and protein N-acetylation, (4) creation of a training set consisting of all 12 samples, which were first subjected to iRT calibration and mProphet peak detection using Skyline, and then further refined by excluding all peptides with dotp values less than 0.8 in all samples (5) enforcing peptide uniqueness by protein and eliminating iRT outliers (6) removing all repeated peptides to represent each protein by unique peptides, and (7) limiting the maximum number of peptides per protein to 10. Applying this stringent set of QC filter steps resulted in a final DIA assay library target list containing 52,361 transitions, 9226 precursors, 9226 peptides, and 2120 proteins (Figure 1). The five most common transitions (fragment ions) included in this kidney DIA assay library are y4 to y8 ions, followed by b4, b5, y3, and y9 ions (Figure 1e). Despite requiring each protein to be represented by at least two peptides in the raw spectral library, one-third (711) of the DIA assay library proteins are only represented by a single peptide (although still by multiple transitions) (Figure 1f). The reason is that at least one peptide was excluded by DIA assay filter steps outlined above because it did not meet one of the QC criteria.

Quantitative proteomics data quality and statistical power for discerning treatment differences

To demonstrate the power of using DIA assay libraries for accurate protein quantitation on a proteome-wide scale we analyzed the DIA data for six replicates of kidneys from fish acclimated to brackish water (BW) and freshwater (FW), each. A volcano plot was generated with Skyline 20.0 as previously described (Li, Levitan, Gomez-Jimenez, & Kültz, 2018) to visualize statistically significant proteins associated with acclimation of fish to BW (Figure 2A). QC charts of the DIA data illustrate that the majority of transitions had a mass error of less than 10 ppm (Figure 2B) and there was a perfect correlation of measured retention time (RT) with the predicted RT based on intrinsic RT (iRT) standards (Escher et al., 2012) with no outliers (Figure 2C). In addition to an adjusted p value < 0.05, which took into account multiple testing (Benjamini & Hochberg, 1995), a fold-change (FC) threshold was enforced in the volcano plots. The FC threshold was 1.85 as calculated by MSstats (Choi et al., 2014) using a statistical power analysis based on the coefficient of variance (CV) of all experimental DIA data (Figure 2D). Furthermore, the vast majority of transition peaks for all 12 samples in this dataset had mProphet (Reiter et al., 2011) peak scores of $q < 0.01$, which was the peak quality threshold for inclusion in MSstats quantitative DIA data analysis (Figure 2e). In this example, 21 kidney proteins were up-regulated and 21 down-regulated during acclimation of fish from FW to BW (Figure 2A).

Parallel quantitation of mRNAs by transcriptomics

In total, 35902 unique transcripts were quantified. The average read length was 80 bp, with a minimum of 5.5 million reads per lane with 4 lanes for each sample. The FASTQ analysis was successful after trimming and adaptors removal for all the sequenced samples.

Analysis for DEG resulted with 17 salinity-dependent transcripts (p-adj < 0.05). From them, 8 transcripts were up-regulated and 9 were down-regulated (Table 2). Within the up-regulated genes, enrichments for

membrane components such as transporters, nucleic acid binding, dehydrogenases/reductases and one immuno-globulin domain-containing protein were present. The down-regulated genes included enrichments notably for regulators of transcription by binding RNA or DNA, signal transduction, endopeptidases for proteolysis and integral components of membrane such as microtubules.

Correlation between protein and corresponding mRNA abundances

Kidney protein and mRNA expression levels had a low level of overall correlation and were also poorly correlated for significantly regulated proteins. Overall, plotting log2 transformed mRNA FC against protein FC resulted in an R^2 value of 0.0061, which indicates lack of any measurable correlation (Figure 3A). The main reason for this result is the almost complete absence of significant regulation of kidney mRNAs during BW acclimation (see previous paragraph). Plotting all significantly up-regulated and down-regulated proteins against the FC of their corresponding mRNA transcripts resulted in a slightly higher R^2 value of 0.185, although this correlation is still weak (Figure 3B). Few of the mRNA transcripts corresponding to significantly regulated proteins had fold changes which met the 1.85 FC threshold. These included two upregulated mRNAs, major facilitator superfamily domain-containing protein 4A (MFSD4A) and *myo* -inositol-3-phosphate synthase 1 (MIPS1), and two downregulated mRNAs, *myo* -inositol oxygenase (MIOX) and dehydrogenase/reductase SDR family member 11 isoform X2 (DHRS11-X2). Despite meeting the FC threshold, none of these four transcripts were statistically significantly different between BW and FW fish and none of the significantly regulated mRNAs matched to genes that are represented as proteins in the DIA assay library.

Expansion of datasets for functional enrichment analyses by clustering

As outlined above, our example dataset generated a relatively small number of significantly regulated proteins, which limits the power of functional enrichment analyses. To expand the set of proteins for such analyses we performed non-biased k-means clustering, which resulted in effectively segregating 37 of the total 42 significantly regulated proteins into one of two clusters, which essentially correspond to up-regulated (cluster 6) and down-regulated (cluster 1) groups of proteins. Cluster 1 was larger and more inclusive, including 20 of the 21 down-regulated proteins and 185 non-significant proteins showing the same overall abundance pattern. Cluster 6 contained 17 of 21 up-regulated proteins and 37 non-significant proteins with a similar overall abundance pattern (Figure 4).

Protein functional network analysis of DIA data

STRING enrichment analysis was performed using abundance data for 2114 detectable proteins out of 2120 proteins included in the DIA assay library. to identify categories that are enriched and depleted following acclimation of fish to BW (Table S1). In addition, a smaller subset of proteins was used for more in-depth network analysis, which included all proteins from clusters 1 and 6, as well as the remaining 5 significantly regulated proteins which were not in clusters 1 or 6. Additionally, all significantly regulated mRNAs were also included by using their corresponding protein accession numbers (ACs) (Figure 5). In total, 277 unique protein ACs were queried with STRING for further network and pathway analysis.

From this list, 174 protein ACs from the DIA assay library and 1 protein AC with a significantly regulated mRNA were mapped to corresponding STRING IDs while the remainder had no STRING ID. The resulting list of 175 STRING IDs was enriched in kidneys of BW fish with FDR<0.05 for "mitochondri-on", "monooxygenase", "oxidoreductase", "NADP", "Microsome", "GTP-binding", "glycolysis", and "FAD" (Table 1). Protein domain enrichment in kidneys of BW fish was most significant for (FDR<0.002) for Al-do/keto reductase, short chain dehydrogenase/reductase SDR, flavin binding monooxygenase, and pyridine nucleotide-disulphide oxidoreductase.

Only using the small set of significantly regulated proteins for STRING analysis yielded many fewer enriched categories than using the complete set of 175 STRING IDs or using sets consisting of all proteins in clusters 1 (down-regulated) and 6 (up-regulated) (Table 1). Significantly regulated proteins also did not correspond to any of the protein identifiers enriched in the whole protein set (Table S1) This result demonstrates the

added value of performing non-biased clustering and STRING analysis in tandem.

When consolidating the list of 175 STRING IDs into a Markov Clustering (MCL) network, 119 IDs were found to have at least one edge between protein nodes included in this list. At an MCL inflation rate of 1.3, proteins were separated into 7 networks that included all 23 significant proteins which were connected to another node by least one edge (Figure 6). Networks 1 and 3 account for 13 of the significant proteins and include 6 of the 8 most highly regulated proteins which were found in the network map, defined as having an FC greater than 4 (Figure 7). Network 1 was enriched for the keyword hydrolase and the protein domains glutathione S-transferase, papain cysteine protease, and thioredoxin-like. Network 3 was enriched for the keyword ligase and the protein domains aldo-keto reductase, acetyl-CoA synthetase, Acetate-CoA ligase, and NADP-dependent oxidoreductase. Other networks were associated with terms associated to the overall list, such as cytochrome C-oxidase for network 4 and short chain dehydrogenase/reductase SDR for network 2 (Table 1), linking STRING networks with specific cellular responses.

In a separate analysis, the query list of 277 unique protein ACs was associated with 236 KEGG orthology (KO) identifiers. From the complete list, over-representation of KO identifiers was greatest for KEGG pathways 01220 (degradation of aromatic compounds), 00625 (Chloroalkane and chloroalkene degradation), and 00982 (Drug metabolism - cytochrome P450). For significantly up-regulated proteins over-representation was greatest in the pathway 00053 (ascorbate and alderate metabolism) due to the presence of UDP-glucoronosyltransferase (UGT) and aldehyde dehydrogenase (ALDH). This pathway was especially important as it also contains the significantly down-regulated MIOX protein (Figure 8). KEGG over-enrichment for the full list can be linked predominantly to the STRING network 3, which is enriched in KEGG pathways 01220, 00625, and 00053. The greatest over-representation of significantly down-regulated proteins was in KEGG pathway 05100 (bacterial invasion of epithelial cells), which was also seen greatly over-represented in network 6 due to the significantly down-regulated proteins actin-related protein 2/3 complex subunit 1A, integrin-linked kinase, and cell division cycle control protein 42, and dynamin-2 isoform X4.

Key kidney proteins associated with BW acclimation of fish were identified based on statistical significance, degree of FC, correlation with mRNA regulation, and their presence in central STRING networks and KEGG pathways. Some of these proteins were represented by only one paralog, but several proteins were found to have one paralog which was significantly regulated, while corresponding paralogs had low-FC or were regulated in the opposite direction. These proteins include the up-regulated Von Willebrand factor A domain-containing (VFA) protein, elongation factor A (EFA), ALDH and UGT, and the down-regulated proteins hemoglobin and DHRS11 (Figure 9).

Discussion

DIA assay libraries enable large-scale ecological proteomics

DIA is a post-acquisition targeted mass spectrometry-based approach for quantitation of thousands of proteins that was first introduced in 2012 (Gillet et al., 2012). This approach is exceptionally well suited for ecological proteomics because it combines the advantages of not requiring any sample pre-treatment or labeling, a robust sample preparation workflow, the ability to internally normalize samples by iRT standards and normalizing to median abundance of all targets with highly accurate and visually traceable quantitation, direct quantitation based on intensity measurements of peptide fragment ions (rather than indirect methods such as those using antibodies), and robust bioinformatics tools for quantitative analyses and public sharing of data (Skyline, MSstats, mProphet, Panorama). Even though proteome coverage in DIA assay libraries is not complete, assaying over 2000 proteins simultaneously and with very high accuracy significantly expands the capabilities of protein quantitation relative to more traditional antibody-based approaches such as Western blots and ELISAs. Moreover, variability due to antibody batch variation, low affinity, or low specificity of antibodies can be avoided when using DIA assay libraries. Furthermore, DIA assay libraries provide a convenient way for normalization of relative protein abundances against the overall sample median and for large-scale comparisons of transcriptome and proteome dynamics in different ecological contexts. This is only possible because thousands of proteins are included in the library and their mean abundance is much more

8

likely to be comparable amongst samples than the abundance of a single or few normalizer proteins such as beta-actin.

The coverage of DIA assay libraries can be increased by pooling proteomes from multiple tissues and multiple purified subcellular compartments, which can yield DIA assay libraries consisting of more than 10,000 proteins (Blattmann et al., 2019). However, we do not prefer pooling because analysis of DIA data for individual tissues using such pooled DIA libraries would generate many missing values, which represents a problem for statistical analysis. In that regard, tissue-specific DIA assay libraries that were generated from representative training samples are preferable. Thus, generation of DIA assay libraries for key tissues of important ecological model organisms breaks new ground for molecular ecology. To illustrate the usefulness of the *O. niloticus* kidney DIA assay library for analyzing molecular phenotypes (proteome dynamics) in different ecological contexts we have analyzed kidneys obtained from fish acclimated to different environmental salinities.

DIA assay libraries enable analyses of mRNA:protein coregulation by ecological contexts

Determining the relationship between mRNA and protein abundance was an important aim of this analysis, which shapes further analytical approaches. This study found very low correlation between mRNA and proteins in kidneys of fish acclimated from FW to BW. The proteins identified as significant using the DIA assay library approach would not have been identified from transcriptomics studies since none of the corresponding mRNAs were significantly regulated. Therefore, functional analyses based on proteome dynamics during salinity stress provide additional rather than merely redundant information when compared with mRNAseq analyses (see below). In theory, lack of coregulation between the mRNA and protein levels of biological organization during salinity acclimation can be attributed to both biological and technical causes. From a technical standpoint, the dynamic range of protein quantity in a tissue sample is several orders of magnitude greater than that for mRNA (Buccitelli & Selbach, 2020). Because the protein dynamic range is so much greater, differences in protein levels can be identified more accurately. Biologically, because protein's action and stability can be modulated by post-translational modification (PTM), cells may not need to alter mRNA levels in order to respond to a stressor (Jovanovic et al., 2015), which has indeed been suggested for salinity stress responses of euryhaline fish (Evans & Somero, 2009). Additionally, proteins typically have a significantly longer half-life than mRNA and are less prone to rapid changes, which reduces the overall variability between replicates for proteins (Schwanhäusser et al., 2011).

DIA assay libraries enable holistic functional analyses of co-regulated clusters

Significance testing is a commonly used method for evaluating proteomic data, but often fails to illuminate the context in which these proteins are functioning. The DIA assay library generated in this study contained over 2000 proteins, which permitted non-biased clustering of all corresponding protein abundance patterns associated with salinity acclimation. In our example, cluster analysis identified groups of proteins which were regulated in similar ways to the significant proteins, capturing 88% of the significant proteins in one of two clusters and expanding the number of proteins to be evaluated by six times while still focusing on proteins responding similarly to a salinity challenge. The value of this expansion was particularly evident in the STRING network analysis, as an analysis of only 42 significant proteins did not return any protein-protein networks with more than one edge. The expanded list returned a complex network indicating the connections between significant proteins with non-significant intermediaries, and the effects of significant regulation on connected proteins despite these effects not reaching the level of statistical significance. Additionally, STRING and KEGG enrichment only returned one protein domain and one KEGG pathway which were highly enriched for significantly up-regulated proteins. Thus, our study is in agreement with previous reports pointing out that systems scale "omics" studies are rendered more powerful when a variety of threads of evidence, including cluster analysis, are used to expand on significance testing results (Gehlenborg et al., 2010).

Regulation of the myo-inositol pathway during salinity stress

Many teleosts, including *O. niloticus* , rely on *myo* -inositol as a compatible osmolyte during salinity stress. The link between *myo* -inositol synthesis and environmental salinity is well established in *Oreochromis* species

(Cramb et al., 2013; Gardell et al., 2013; Kalujnaia, McVee, Kasciukovic, Stewart, & Cramb, 2010; Sacchi, Gardell, Chang, & Kültz, 2014; Sacchi, Li, Villarreal, Gardell, & Kültz, 2013). *Myo* -inositol is an organic osmolyte which is used to maintain osmotic pressure within the cell to overcome the negative effects of small and charged inorganic ions. Our data are in agreement with the previous literature as they indicate up-regulation of renal *myo* -inositol-3-phospate synthase (MIPS) during BW acclimation of fish (FC=3.35, p=.042). Interestingly, in *O. mossambicus myo* -inositol production is consistently increased through up-regulation of IMPase1.1, which is not differentially regulated in *O. niloticus* kidney (FC=1.001, p=.99, Supplementary Table 2). On the other hand, one of the most highly significant proteins in *O. niloticus* kidney is down-regulated *myo* -inositol oxygenase (MIOX, FC=-6.22, p=.0058), which is a key enzyme for the degradation of *myo* -inositol. MIOX has been shown to be important in determining *myo* -inositol levels *in vivo* (Arner et al., 2001). Transcripts of MIOX have been found to be differentially regulated in response to salinity acclimation in ayu (*Plecoglossus altivelis*) larvae (Lu, Zhang, Yang, Li, & Chen, 2016) and turbot (*Scophthalmus maximus*) (Cui et al., 2020). Nonetheless, MIOX regulation has not been previously noted in *Oreochromis* species.

The most highly regulated protein in our study is major facilitator superfamily domain-containing protein 4A (MFSD4A) (FC=8.44, p=.001), a transmembrane glucose transporter. Glucose, in addition to being an important source of energy, is the substrate of MIPS (in the form of glucose 6-phosphate), indicating that upregulation of MFSD4A facilitates*myo* -inositol production. Another indication that MFSD4A is linked to *myo* -inositol production is that it is one of the few proteins that has good correlation between protein and mRNA regulation, along with the other important *myo* -inositol related proteins MIPS and MIOX (Figure 3).

Our network and pathway analysis of DIA assay data demonstrates that MIOX is connected to many of the other highly salinity-regulated proteins. The most highly enriched KEGG pathway in BW acclimated fish containing significantly regulated proteins is the ascorbate and alderate metabolism pathway. This finding indicates the compensatory upregulation of UGT and ALDH enzymes to replace the products of*myo* -inositol degradation by MIOX. MIOX and ALDH3 (FC=4.65, p=.024) were both found in STRING network 3, and MIOX was also connected to a (non-significant) isoform of UGT in network 2 (Figure 6). ALDH and UGT also function in antioxidant metabolism and detoxification of xenobiotics during environmental pollution.

Regulation of antioxidant functions during salinity stress

Aldehydes are toxic carbonyl compounds which can be created in the body as a byproduct of lipid peroxidation and a variety of other metabolic processes (Laskar & Younus, 2019), and are converted into non-harmful compounds by ALDH. In one study, lipid peroxidation increased significantly by 48h after initially decreasing following acute transfer to high salinity in the croaker fish *Pseudosciaena crocea* , suggesting that antioxidant pools were reduced and that the system was overwhelmed over time (Zeng, Ai, Wang, Zhang, & Wu, 2017). Other regulated cellular functions align well with known responses to salinity stress including additional solute transporters in network 2, oxidative phosphorylation in network 4, and DNA protection in network 5. These findings agree with the notion that oxidative stress is a secondary, non-specific consequence of osmotic and other environmental stresses, resulting from upregulation of oxidative metabolism and the associated increase in oxygen radical leakage (Kültz, 2020). ALDH3 is up-regulated in response to hydrocarbon exposure in mice (Lindros et al., 1998), and found to be highly responsive to phenylurea herbicides in teleost fish (Marlatt & Martyniuk, 2017). The specific isoform of UGT significantly up-regulated in this data set, UGT2C1-X2 (FC=4.03, p=.030) , was found to be transcriptionally up-regulated in *Cynoglossus semilaevis*flatfish in response to the environmental pollutant Perfluorooctane sulfonate (Zhang, Sun, Chen, Zhang, & Cai, 2020).

Many of the other proteins found in network 3 have detoxification/ antioxidant functions, including the aldo-keto reductases and the significantly regulated quinoid dihydropteridine reductase (QDPR) (FC=9.84, p=.039). QDPR was shown to induce production of major antioxidant enzymes including glutathione peroxidase 3 and superoxide dismutase 1 in cell culture (Gu et al., 2017). Like ALDH and DHRS (significant but not found in STRING network 3), aldo/keto reductases are another system for metabolizing toxic aldehydes

(Laskar & Younus, 2019) but they were non-significantly down-regulated in BW acclimated tilapia. Therefore, our data suggest that ALDH3 and QDPR are more important for kidney function in BW acclimated tilapia than the aldo/keto reductases.

Many of the proteins in STRING network 1 are also involved in protecting against oxidative stress, including thioredoxin related proteins and several glutathione S-transferase (GST) isoforms. Evidence from rats exposed to polycyclic hydrocarbons indicates co-induction of ALDH3 along with GST in liver tissue, which points to common enhancers or transcription factors (Lindros et al., 1998). Three different GST isoforms are connected in network 1 to another highly regulated protein, elongation factor 1-delta (FC=4.48, p=.0025). EF 1-delta was found to be regulated in several studies of fish exposed to pollutants (Jeffries, Brander, Britton, Fangue, & Connon, 2015; Williams, Gensberg, Minchin, & Chipman, 2003) and putative disease agents (Lü et al., 2014), as well as hypo-osmotic stress in shrimp (Liu et al., 2016).

Coordination of intra- and extracellular signaling during salinity stress

Von Willebrand factor A domain-containing (VFA) proteins are extracellular proteins involved in cell-cell adhesion and signaling as well as blood clot formation after injury. VFA domain-containing protein 8 (VFA8) was the third most highly regulated protein (FC=6.41, p=.0001) while VFA/sushi/EGF and pentraxin domain-containing protein 1 was the second most significantly regulated mRNA (FC=-17.4, p=1.02E-6). This opposite regulation pattern suggests that different VFA proteins are associated with renal function in FW versus BW acclimated tilapia. In one study, red blood cell count was significantly reduced in *O. niloticus* exposed to 12ppt saline water for 14 days (Elarabany, Bahnasawy, Edrees, & Alkazagli, 2017). Osmotic stress is known to induce erythrocyte cell death by opening Ca^{2+}-permeable cation channels, increasing cytosolic Ca^{2+} activity and triggering erythrocyte apoptosis (Lang et al., 2003). Therefore, BW acclimation may increase erythrocyte cell death in the kidney. This conclusion is supported by the significant reduction in hemoglobin subunit beta-1 (FC=-7.27, p=.0051) and increase in haptoglobin (FC=2.56, p=.03), which neutralizes the oxidative effects of free heme groups. Furthermore, VFA8 has been reported to translocate to the inner membrane of mitochondria in response to calcium-induced signaling and to influence mitochondrial energy metabolism (Luo et al., 2020). These data suggest that VFA proteins are involved in linking extracellular osmotic stress sensing with intracellular responses such as increased mitochondrial energy production.

Conclusions

Understanding how organisms respond to environmental stresses and other ecological contexts requires analyzing multiple levels of biological organization. Here we introduce an approach to molecular ecology that allows targeted quantitation of proteins on a proteome-wide scale by employing DIA assay libraries. We demonstrate the utility of this approach by comparing kidney proteomes of *O. niloticus* exposed to different salinities. Our study shows that regulation at the proteome greatly differs from that at the transcriptome level. Therefore, analyses of these two levels of biological organization produce different types of knowledge on tissue-specific phenotypes and functions that are associated with salinity acclimation. Moreover, analyzing the mRNA:protein regulation ratio provides insight into mechanisms of regulation (transcriptional, post-transcriptional, translational, post-translational). The DIA assay library generated in this study enabled a global picture of kidney proteome dynamics during salinity stress. Clustering significantly regulated proteins with other proteins that had similar (but non-significant) patterns of regulation, expanded the data set and revealed interconnectedness between known and novel salinity stress responses. The role of the osmolyte *myo*-inositol as an important cellular protective molecule was highlighted and *myo*-inositol degradation identified as an important mechanism in *O. niloticus* kidney. Cellular protection and detoxification by antioxidant mechanisms was another core function associated with BW acclimation of *O. niloticus*, which was revealed using the kidney DIA assay library. Another interesting finding was the potential of red blood cell damage to be the trigger for the increase in oxidative metabolism via the signaling molecule VFA.

In summary, this study introduces the DIA assay library approach as a novel tool for proteome-wide molecular ecology studies. The utility of DIA assay libraries for discerning different ecological contexts was

11

demonstrated by using a kidney DIA assay library to analyze proteome dynamics in response to acclimation of *O. niloticus* from FW to BW and by highlighting large differences in the regulation at the proteome versus transcriptome levels of biological organization.

References

Anders S., Huber W. (2010). Differential expression analysis for sequence count data. *Genome Biology* , 11, R106. doi: 10.1186/gb-2010-11-10-r106, http://genomebiology.com/2010/11/10/R106/.

Anders S., Pyl P.T., Huber W. (2014). HTSeq – A Python framework to work with high-throughput sequencing data. Bioinformatics doi: 10.1093/bioinformatics/btu638.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arner, R. J., Prabhu, K. S., Thompson, J. T., Hildenbrandt, G. R., Liken, A. D., & Reddy, C. C. (2001). myo-Inositol oxygenase: molecular cloning and expression of a unique enzyme that oxidizes myo-inositol and D-chiro-inositol. *Biochemical Journal* , *360* (Pt 2), 313–320.

Arnhard, K., Gottschall, A., Pitterl, F., & Oberacher, H. (2015). Applying "Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra" (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry. *Anal Bioanal Chem* , *407* (2), 405–414. doi: 10.1007/s00216-014-8262-1

Batchelor, E., Loewer, A., Mock, C., & Lahav, G. (2011). Stimulus-dependent dynamics of p53 in single cells. *Mol Syst Biol* , *7* , 488. doi: 10.1038/msb.2011.20

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* , *57* (1), 289–300.

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics* , *15* (3), 163–175. doi: 10.1038/nrg3662

Blattmann, P., Stutz, V., Lizzo, G., Richard, J., Gut, P., & Aebersold, R. (2019). Generation of a zebrafish SWATH-MS spectral library to quantify 10,000 proteins. *Scientific Data* , *6* , 190011. doi: 10.1038/sdata.2019.11

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

Buccitelli, C., & Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics* , 1–15. doi: 10.1038/s41576-020-0258-4

Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., & Vitek, O. (2014). MS-stats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* , *30* (17), 2524–2526. doi: 10.1093/bioinformatics/btu305

Clarke, B. (1971). Natural Selection and the Evolution of Proteins. *Nature* , *232* (5311), 487. doi: 10.1038/232487a0

Cramb, G., Kalujnaia, S., Gellatly, S., Hazon, N., Villasenor, A., & Yancey, P. (2013). Expression and functions of inositol monophosphatase (IMPA) in seawater (SW)-acclimated euryhaline teleosts. *The FASEB Journal* , *27* (S1), 937.7-937.7. doi: 10.1096/fasebj.27.1_supplement.937.7

Crowgey, E. L., Matlock, A., Venkatraman, V., Fert-Bober, J., & Van Eyk, J. E. (2017). Mapping biological networks from quantitative data-independent acquisition mass spectrometry: data to knowledge pipelines. *Methods Mol Biol* , *1558* , 395–413. doi: 10.1007/978-1-4939-6783-4_19

Cui, W., Ma, A., Huang, Z., Wang, X., Liu, Z., Xia, D., . . . Zhao, T. (2020). Comparative transcriptomic analysis reveals mechanisms of divergence in osmotic regulation of the turbot Scophthalmus maximus. *Fish Physiology and Biochemistry* , *46* (4), 1519–1536. doi: 10.1007/s10695-020-00808-6

Dobin A, Davis CA, Schlesinger F, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* , 29(1):15-21. doi:10.1093/bioinformatics/bts635

Ebhardt, H. A., Root, A., Sander, C., & Aebersold, R. (2015). Applications of targeted proteomics in systems biology and translational medicine. *Proteomics* , *15* (18), 3193–3208. doi: 10.1002/pmic.201500004

Elarabany, N., Bahnasawy, M., Edrees, G., & Alkazagli, R. (2017). Effects of Salinity on Some Haematological and Biochemical Parameters in Nile Tilapia, *Oreochromus niloticus* . *Agriculture, Forestry and Fisheries* , *6* (6), 200. doi: 10.11648/j.aff.20170606.13

Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., . . . Rinner, O. (2012). Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* , *12* (8), 1111–1121. doi: 10.1002/pmic.201100463

Evans, T. G. (2015). Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. *JOURNAL OF EXPERIMENTAL BIOLOGY* , *218* (12, SI), 1925–1935. doi: 10.1242/jeb.114306

Evans, T. G., & Somero, G. N. (2009). Protein-protein interactions enable rapid adaptive response to osmotic stress in fish gills. *Communicative & Integrative Biology* , *2* (2), 94–96.

Franks, A., Airoldi, E., & Slavov, N. (2017). Post-transcriptional regulation across human tissues. *PLOS Computational Biology* , *13* (5), e1005535. doi: 10.1371/journal.pcbi.1005535

Frömmel, C., & Holzhütter, H.-G. (1985). An estimate on the effect of point mutation and natural selection on the rate of amino acid replacement in proteins. *Journal of Molecular Evolution* , *21* (3), 233–257. doi: 10.1007/BF02102357

Gardell, A. M., Yang, J., Sacchi, R., Fangue, N. A., Hammock, B. D., & Kültz, D. (2013). Tilapia (Oreochromis mossambicus) brain cells respond to hyperosmotic challenge by inducing myo-inositol biosynthesis. *The Journal of Experimental Biology* , *216* (Pt 24), 4615–4625. doi: 10.1242/jeb.088906

Gehlenborg, N., O'Donoghue, S. I., Baliga, N. S., Goesmann, A., Hibbs, M. A., Kitano, H., . . . Gavin, A.-C. (2010). Visualization of omics data for systems biology. *Nature Methods* , *7* (3 Suppl), S56-68. doi: 10.1038/nmeth.1436

Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., . . . Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* , *11* (6), O111 016717. doi: 10.1074/mcp.O111.016717

Gu, Y., Wang, Y., Zhang, H., Zhao, T., Sun, S., Wang, H., . . . Li, P. (2017). Protective effect of dihydropteridine reductase against oxidative stress is abolished with A278C mutation. *Journal of Zhejiang University. Science. B* , *18* (9), 770–777. doi: 10.1631/jzus.B1600123

Huang, Q., Yang, L., Luo, J., Guo, L., Wang, Z., Yang, X., . . . Zhang, Y. (2015). SWATH enables precise label-free quantification on proteome scale. *Proteomics* , *15* (7), 1215–1223. doi: 10.1002/pmic.201400270

Jeffries, K. M., Brander, S. M., Britton, M. T., Fangue, N. A., & Connon, R. E. (2015). Chronic exposures to low and high concentrations of ibuprofen elicit different gene response patterns in a euryhaline fish. *Environmental Science and Pollution Research International* ,*22* (22), 17397–17413. doi: 10.1007/s11356-015-4227-y

Jovanovic, M., Rooney, M. S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., . . . Regev, A. (2015). Dynamic profiling of the protein life cycle in response to pathogens. *Science* , *347* (6226). doi: 10.1126/science.1259038

Kalujnaia, S., McVee, J., Kasciukovic, T., Stewart, A. J., & Cramb, G. (2010). A role for inositol monophosphatase 1 (IMPA1) in salinity adaptation in the euryhaline eel (Anguilla anguilla). *Faseb Journal* , *24* (10), 3981–3991. doi: 10.1096/fj.10-161000

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* ,*428* (4), 726–731. doi: 10.1016/j.jmb.2015.11.006

Keerthikumar, S., & Mathivanan, S. (2017). Proteotypic Peptides and Their Applications. *Methods in Molecular Biology (Clifton, N.J.)* ,*1549* , 101–107. doi: 10.1007/978-1-4939-6740-7_8

Koopmans, F., Ho, J. T. C., Smit, A. B., & Li, K. W. (2018). Comparative analyses of data independent acquisition mass spectrometric approaches: DIA, WiSIM-DIA, and untargeted DIA. *Proteomics* ,*18* (1), 1–6. doi: 10.1002/pmic.201700304

Kratochwil, C. F., & Meyer, A. (2015). Closing the genotype-phenotype gap: emerging technologies for evolutionary genetics in ecological model vertebrate systems. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* , *37* (2), 213–226. doi: 10.1002/bies.201400142

Kültz, D. (2020). Evolution of cellular stress response mechanisms.*Journal of Experimental Zoology Part A: Ecological and Integrative Physiology* , *n/a* (n/a). doi: 10.1002/jez.2347

Kültz, D., Li, J., Gardell, A., & Sacchi, R. (2013). Quantitative molecular phenotyping of gill remodeling in a cichlid fish responding to salinity stress. *Molecular & Cellular Proteomics* , *12* (12), 3962–3975. doi: 10.1074/mcp.M113.029827

Lang, K. S., Duranton, C., Poehlmann, H., Myssina, S., Bauer, C., Lang, F., . . . Huber, S. M. (2003). Cation channels trigger apoptotic death of erythrocytes. *Cell Death and Differentiation* ,*10* (2), 249–256. doi: 10.1038/sj.cdd.4401144

Laskar, A. A., & Younus, H. (2019). Aldehyde toxicity and metabolism: the role of aldehyde dehydrogenases in detoxification, drug resistance and carcinogenesis. *Drug Metabolism Reviews* , *51* (1), 42–64. doi: 10.1080/03602532.2018.1555587

Li, J., Levitan, B., Gomez-Jimenez, S., & Kültz, D. (2018). Development of a Gill Assay Library for Ecological Proteomics of Threespine Sticklebacks (Gasterosteus aculeatus). *Molecular & Cellular Proteomics: MCP* , *17* (11), 2146–2163. doi: 10.1074/mcp.RA118.000973

Lindros, K. O., Oinonen, T., Kettunen, E., Sippel, H., Muro-Lupori, C., & Koivusalo, M. (1998). Aryl hydrocarbon receptor-associated genes in rat liver: Regional coinduction of aldehyde dehydrogenase 3 and glutathione transferase Ya. *Biochemical Pharmacology* ,*55* (4), 413–421. doi: 10.1016/S0006-2952(97)00495-4

Liu, H., Sun, W., Dong, X., Chi, S., Yang, Q., Li, Y., & Tan, B. (2016). Profiling of Up-Regulated Genes Response to Acute Hypo-Osmotic Stress in Hepatopancreas and Gill of the Pacific White Shrimps (Litopenaeus vannamei). *International Journal of Biology* ,*8* (2), p43. doi: 10.5539/ijb.v8n2p43

Lü, A., Hu, X., Wang, Y., Shen, X., Li, X., Zhu, A., . . . Feng, Z. (2014). iTRAQ analysis of gill proteins from the zebrafish (Danio rerio) infected with Aeromonas hydrophila. *Fish & Shellfish Immunology* ,*36* (1), 229–239. doi: 10.1016/j.fsi.2013.11.007

Lu, X.-J., Zhang, H., Yang, G.-J., Li, M.-Y., & Chen, J. (2016). Comparative transcriptome analysis on the alteration of gene expression in ayu (Plecoglossus altivelis) larvae associated with salinity change.*Zoological Research* , *37* (3), 126–135. doi: 10.13918/j.issn.2095-8137.2016.3.126

Luo, M., Ma, W., Sand, Z., Finlayson, J., Wang, T., Brinton, R. D., . . . Mandarino, L. J. (2020). Von Willebrand factor A domain-containing protein 8 (VWA8) localizes to the matrix side of the inner mitochondrial membrane. *Biochemical and Biophysical Research Communications* , *521* (1), 158–163. doi: 10.1016/j.bbrc.2019.10.095

Marlatt, V. L., & Martyniuk, C. J. (2017). Biological responses to phenylurea herbicides in fish and amphibians: New directions for characterizing mechanisms of toxicity. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* , *194* , 9–21. doi: 10.1016/j.cbpc.2017.01.002

Mularoni, L., Ledda, A., Toll-Riera, M., & Albà, M. M. (2010). Natural selection drives the accumulation of amino acid tandem repeats in human proteins. *Genome Research* , *20* (6), 745–754. doi: 10.1101/gr.101261.109

Pino, L. K., Searle, B. C., Bollinger, J. G., Nunn, B., MacLean, B., & MacCoss, M. J. (2017). The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev* ,*39* (3), 229–244. doi: 10.1002/mas.21540

Pohl, C., & Dikic, I. (2019). Cellular quality control by the ubiquitin-proteasome system and autophagy. *Science* ,*366* (6467), 818–822. doi: 10.1126/science.aax3769

R Core Team (2020). R: A language and Environment for Statistical Computing. *R foundation for Statistical Computing* , Vienna, Austria. URL https://www.R-project.org

Reiter, L., Rinner, O., Picotti, P., Huttenhain, R., Beck, M., Brusniak, M. Y., . . . Aebersold, R. (2011). mProphet: automated data processing and statistical validation for large-scale SRM experiments.*Nat Methods* , *8* (5), 430–435. doi: 10.1038/nmeth.1584

Sacchi, R., Gardell, A. M., Chang, N., & Kültz, D. (2014). Osmotic regulation and tissue localization of the myo-inositol biosynthesis pathway in tilapia (Oreochromis mossambicus) larvae. *Journal of Experimental Zoology. Part A, Ecological Genetics and Physiology* ,*321* (8), 457–466. doi: 10.1002/jez.1878

Sacchi, R., Li, J., Villarreal, F., Gardell, A. M., & Kültz, D. (2013). Salinity-induced regulation of the myo-inositol biosynthesis pathway in tilapia gill epithelium. *The Journal of Experimental Biology* ,*216* (Pt 24), 4626–4638. doi: 10.1242/jeb.093823

Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., . . . Aebersold, R. (2015). Building high-quality assay libraries for targeted analysis of SWATH MS data.*Nat Protoc* , *10* (3), 426–441. doi: 10.1038/nprot.2015.015

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., . . . Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* , *473* (7347), 337–342. doi: 10.1038/nature10098

Schwenk, K., Padilla, D. K., Bakken, G. S., & Full, R. J. (2009). Grand challenges in organismal biology. *Integrative and Comparative Biology* , *49* (1), 7–14. doi: 10.1093/icb/icp034

Stillman, J. H., Denny, M., Padilla, D. K., Wake, M. H., Patek, S., & Tsukimura, B. (2011). Grand opportunities: strategies for addressing grand challenges in organismal animal biology. *Integrative and Comparative Biology* , *51* (1), 7–13. doi: 10.1093/icb/icr052

Sturn, A., Quackenbush, J., & Trajanoski, Z. (2002). Genesis: cluster analysis of microarray data. *Bioinformatics* , *18* (1), 207–208.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., . . . Mering, C. von. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* , *47* (D1), D607–D613. doi: 10.1093/nar/gky1131

Tahmasebi, S., Khoutorsky, A., Mathews, M. B., & Sonenberg, N. (2018). Translation deregulation in human disease. *Nature Reviews Molecular Cell Biology* , *19* (12), 791–807. doi: 10.1038/s41580-018-0034-x

Vowinckel, J., Capuano, F., Campbell, K., Deery, M. J., Lilley, K. S., & Ralser, M. (2013). The beauty of being (label)-free: sample preparation methods for SWATH-MS and next-generation targeted proteomics. *F1000Res* , *2* , 272. doi: 10.12688/f1000research.2-272.v2

Williams, T. D., Gensberg, K., Minchin, S. D., & Chipman, J. K. (2003). A DNA expression array to detect toxic stress response in European flounder (Platichthys flesus). *Aquatic Toxicology (Amsterdam, Netherlands)* , *65* (2), 141–157. doi: 10.1016/s0166-445x(03)00119-x

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics* , *14* (7), 507–515. doi: 10.1038/nrg3457

Zeng, L., Ai, C.-X., Wang, Y.-H., Zhang, J.-S., & Wu, C.-W. (2017). Abrupt salinity stress induces oxidative stress via the Nrf2-Keap1 signaling pathway in large yellow croaker Pseudosciaena crocea. *Fish Physiology and Biochemistry* , *43* (4), 955–964. doi: 10.1007/s10695-016-0334-z

Zhang, L., Sun, W., Chen, H., Zhang, Z., & Cai, W. (2020). Transcriptomic Changes in Liver of Juvenile Cynoglossus semilaevis following Perfluorooctane Sulfonate Exposure. *Environmental Toxicology and Chemistry* , *39* (3), 556–564. doi: 10.1002/etc.4633

Tables and Figures

Table 1: STRING and KEGG enrichment. Heat map of STRING IDs from list of 175 queries (174 proteins and 1 mRNA) which are significantly enriched (FDR<.05) for keywords (green) assigned by Uniprot and protein domains assigned (orange) by Pfam, InterPro, and SMART databases. KEGG IDs (purple) with over-representation >5, defined as the proportion of KEGG IDs in a KEGG pathway versus total KEGG IDs in the list of 236 queries (229 proteins and 7 mRNA) divided by the proportion of KEGG IDs in a pathway from the complete list of 2114 proteins in the DIA assay library versus the total number of KEGG IDs in the DIA assay library. Additionally, individual subsets of the query list (sig. up, sig. down, cluster 1, cluster 6, individual STRING networks) shown in each column and analyzed in the same manner. Darker color indicates higher significance, with FDR (STRING) and over-representation (KEGG) shown for each term.

Figure 1: DIA assay library properties relative to raw spectral library. DIA assay library of O. niloticus kidney. The initial spectral library (SL) represents over 7,000 proteins, 90,000 precursors, and 450,000 million transitions. Seven QC filters were applied to create the DIA assay library containing 2120 proteins (A), 9226 peptides (B), 9226 precursors (C), and 52,361 transitions (D). Most proteins are represented by at least 2 diagnostic peptides. The remainder (25%) was identified by at least 2 unique peptides but only 1 remains after applying all DIA QC filters. The initial bar labeled SL depicts data for the raw spectral library and final bar (step 7) target depicts data for the DIA assay library. Library filtration steps one to six are explained in the text. E , Frequency distributions of fragment ion types represented in the final DIA assay library. F , Frequency distribution for the number of peptides per protein in the DIA assay library. The data were generated with Skyline 20.0 (MacCoss Lab., University of Washington).

Figure 2: Volcano plot and QC of kidney DIA assay data from FW and BW acclimated fish. A , Volcano plot comparing relative protein abundances in FW and BW acclimated fish. Blue circles represent proteins that are significantly reduced on BW fish at adjusted p < 0.05 and FC [?] 1.85. Orange circles represent proteins that are significantly elevated on BW at adjusted p < 0.05 and FC [?] 1.85. Gray circles represent non-significant proteins. B , Mass error distribution of all transitions. C , Retention time (RT)

reproducibility of all peptides in the DIA assay library based on internal RT (iRT) standards. D , Fold change (FC) and coefficient of variation (CV) depending on number of biological replicates at a statistical power of 0.8 and false discovery rate (FDR) of 0.05. E , mProphet peak score distribution indicating that the great majority of peaks meet the threshold for inclusion in MSstats quantitative DIA data (q < 0.01, MacCoss Lab., University of Washington).

Figure 3: Correlation between protein and mRNA regulation. Correlation between difference in mRNA abundance and protein abundance between treatments expressed as -log2 of fold change (FC). Correlation between all 2114 proteins in the DIA assay library and corresponding mRNA transcripts (left) and correlation between all 42 significant proteins defined as p<0.05 and FC>1.85 (right).

Figure 4: Genesis clustering of protein expression. Non-biased clustering of all proteins in the DIA assay. Patterns of protein abundance within replicates was clustered using K-means method. Proteins were clustered into 6 groups with most significant proteins clustered in cluster 1 (down-regulated) and cluster 6 (up-regulated) (left). Individual protein patterns and cluster average indicate the general regulation patterns within clusters, with replicate identifiers (right)

Figure 5: Visual depiction of analysis following Genesis clustering. Of total 2114 proteins in the DIA assay library, 263 were analyzed using KEGG and STRING online tools. These were: 205 proteins in cluster 1, 53 proteins in cluster 6, and 5 additional significant proteins which were not in clusters 1 & 6. Of these proteins, 174 had unique STRING identifiers and 229 had KEDD IDs. Additionally, 14 significantly regulated mRNA were added to the query lists, of which 1 matched a STRING ID and 7 had matching KEGG IDs.

Figure 6: STRING network map containing all networks with at least one significantly regulated protein (p<.05, FC>1.85). Nodes represent one STRING protein ID and edges represent a connection based on protein-protein interactions including known or predicted interactions based on published literature. The network map is split into 7 different networks based on score of combined connections, with a Markov Cluster (MCL) inflation factor of 1.3. Large circle bolding represents proteins with greater than 4-fold difference between treatments, while smaller bolding represents proteins with significant regulation and between 1.85 and 4-fold difference between treatments.

Figure 7: Expansion of STRING networks 1 and 3 from the full network map. Blue circles represent significantly downregulated proteins while orange circles represent significantly upregulated proteins. The legend indicates the fold change difference between treatments, with the largest circle representing greater than 5-fold difference, the next largest representing between 3 and 5-fold difference, the next largest representing between 1.85 and 3-fold difference, and the smallest bolded circle representing proteins which were statistically significant (p<.05) but which did not meet the fold change cut-off (FC>1.85).

Figure 8: Ascorbate and alderate metabolism KEGG pathway. Significant proteins found in the KEGG pathway of ascorbate and alderate metabolism (map00053). Dark orange indicates significant upregulation, dark blue indicates significant downregulation, and light blue indicates non-significant downregulation. Pathway is taken from portion of KEGG pathway provided by Kanehisa Laboratories 11/9/2018.

Figure 9: Protein isoform comparison. Volcano plots of all proteins and sets of isoforms for several important proteins. All proteins in DIA assay library are shown with log2 transformed Fold Change along the X-axis and log10 transformed p-values along the Y-axis (A). Orange points in the upper right represent significantly upregulated proteins and blue points in the upper left represent significantly downregulated proteins. Dashed lines represent p-value cutoff (p<.05, horizontal line) and FC cutoff (FC>+-1.85, vertical lines). Individual proteins shown include one significant isoform and related non-significant isoforms.

Additional Files

Supporting/Supplemental Information

Supplementary Table S1: STRING enrichment results using log2 FC data for all 2114 proteins included in

the DIA assay library, including enriched STRING clusters, Uniprot keywords and protein domains classified by Pfam and InterPro.

Supplementary Table S2 : Protein accession number and annotation of all proteins included in the *O. niloticus* kidney DIA assay library. Fold-change (BW vs. FW) and adjusted p values are included for the experiment discussed in the text. Data were generated with Skyline 20.0

STRING and KEGG enrichment

	All	Sig Up	Sig Down	Cluster 1	Cluster 6	Network 1	Network 2	Network 3	Network 4	Network 5	Network 6	Network 7	
oxidoreductase	3.0e-02				5.6e-06		4.0e-02						keyword
NADP	4.2e-02				4.8e-04								
monooxygenase	1.6e-02				5.6e-06		4.0e-02						
mitochondrion	1.3e-02		1.9e-03							2.3e-05			
microsome	4.2e-02				7.2e-04								
ligase								3.0e-05					
hydrolase									2.8e-04				
GTP-binding											1.7e-05		
glycosyl transferase	4.2e-02		2.4e-02	1.4e-02									
glycolysis	4.2e-02		2.8e-02				2.2e-04						
FAD	4.2e-02				1.4e-03								
chromosome										4.3e-05			
UDP-glycosyltransferase	5.0e-02				6.5e-03								protein domain
thioredoxin-like	1.5e-02		5.5e-03			2.8e-04							
stanniocalcin	1.5e-02				4.5e-04						4.2e-06		
Small GTPase Rho	4.5e-02						3.5e-03						
short chain dehydrogenase/reductase SDR	1.3e-03				3.6e-05								
pyradine nucleotide-disulphide oxidoreductase	8.3e-04											4.6e-05	
phosphoribosyl transferase					4.5e-02								
peptidase C1A	3.6e-03				1.6e-03	1.9e-04							
papain family cysteine protease	4.0e-03				1.9e-03	6.1e-05							
NADP-dependent oxidoreductase	2.1e-03				7.8e-04			6.6e-08					
NAD(P)-binding	2.1e-03	1.3e-02			6.5e-03		2.6e-02	6.9e-03					
mitochondrial carrier		2.0e-02											
L-lysine 6-monooxygenase	4.0e-03				2.6e-05								
histone 2A	9.0e-03				5.0e-03					9.9e-07			
glutathione S-transferase	8.9e-03				6.4e-03	6.1e-05							
flavin-binding monooxygenase	1.4e-03				1.3e-05								
enoyl-(acyl carrier carrier) reductase	4.1e-03						5.5e-03						
cytochrome C-oxidase	1.3e-02				9.9e-03				4.7e-05				
cysteine peptidase	2.1e-03				7.8e-04	1.9e-04							
amidase	1.5e-02	2.9e-04											
aldo/keto reductase	1.3e-03				3.6e-04				2.9e-08				
Acetyl-coenzyme A synthetase	2.4e-02								2.6e-04				
Acetate-CoA ligase	1.8e-02								1.7e-04				
Steroid hormone biosynthesis				18.4									KEGG pathway
Purine metabolism						22.4						23.1	
Proximal tubule bicarbonate reclamation													
PPAR signaling pathway				5.4									
Ovarian steroidogenesis				21.5									
ntifolate resistance			6.1										
Non-alcoholic fatty liver disease								23.6					
nitrogen metabolism	5.1												
methane metabolism	5.9		7.9			26.2	19.1						
metabolism of xenobiotics by cytochrome P450	5.4												
glycosphingolipid biosynthesis – globo and isoglobo series	5.7		5.1		31.1								
Glycerolipid metabolism									28.7				
Fc gamma R-mediated phagocytosis											35.4		
drug metabolism – cytochrome P450	7.6		6.1	17.2									
degradation of aromatic compounds	7.6		10.2					86.0					
cyanoamino acid metabolism	5.1		6.8										
chloroalkane and chloroalkene degradation	7.6							86.0					
Cardiac muscle contraction									25.3				
Caprolactam degradation									34.4				
Biosynthesis of amino acids			6.7				8.7						
bile secretion	5.1			19.1		17.4							
Bacterial invasion of epithelial cells			7.0								30.9		
Axon guidance											26.8		
ascorbate and alderate metabolism		23.4		12.3			36.9						
antifolate resistance	6.1												

All proteins

Significant proteins

major facilitator superfamily
domain−containing protein 4A

inositol−3−phosphate
synthase 1

myo−inositol oxygenase

$y = -0.026 + 0.11\,x$
$R^2 = 0.0062$

$y = 0.2 + 0.18\,x$
$R^2 = 0.185$

log2(FC) mRNA

log2(FC) protein

myo-Inositol

UDP-glucuronosyltransferase

Myo-inositol oxygenase

1.13.99.1

2.4.1.17

UDP-D-glucose

1.1.1.22

UDP-
D-glucuronate

2.7.7.44

2.7.7.64

D-Glucuronate-1P

2.7.1.43

D-Glucurono-
lactone

Aldehyde dehydrogenase

1.2.1.3

3.1.1.19

D-Glucuronate

D-Glucarate

1.1.1.19

1.1.1.20

L-Gulose-1P

L-Gulose

L-xylo-Hexulono-
lactone

1.1.3.8

3.1.1.17

L-Gulonate

3.1.1.25

L-Gulono-
1,4-lactone

Regucalcin

Significant down   None   Significant up

Aldehyde dehydrogenase

Elongation factor 1

Dehydrogenase/reductase SDR
family member 11

delta−X2

X2

3−A1

9−A1

9−A3−B

8−A1   9−A1

alpha

gamma

delta−X4   epsilon

alpha 1

X4   X5

UDP-glucuronosyltransferase

Von Willebrand factor A domain

Hemoglobin subunit

2C1 − X2

8

beta−1

2C1   2A2 − X3

2C1 − X1   2A2

5A   5A − X1
5A − X1
5A − X1
5A

alpha
(embryonic)

alpha−A   alpha

−LOG10(adj.p)

log2(FC)